

04CO

3/13/01  
OIPE#2  
b4/5

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/782,714

DATE: 03/29/2001  
TIME: 09:59:32

Input Set : N:\CrF3\RULE60\09782714.txt  
Output Set: N:\CRF3\03292001\I782714.raw

3 <110> APPLICANT: Darst, Seth A  
4       Zhang, Gongyi  
5       Campbell, Elizabeth  
6       Minakin, Leonid  
7       Severinov, Konstantin  
9 <120> TITLE OF INVENTION: A CRYSTAL OF BACTERIAL CORE RNA POLYMERASE AND METHODS  
10      OF USE THEREOF  
12 <130> FILE REFERENCE: 600-1-258  
14 <140> CURRENT APPLICATION NUMBER: 09/782,714  
15 <141> CURRENT FILING DATE: 2001-02-13  
17 <150> PRIOR APPLICATION NUMBER: 09/396,651  
18 <151> PRIOR FILING DATE: 1999-09-15  
21 <160> NUMBER OF SEQ ID NOS: 4  
23 <170> SOFTWARE: PatentIn Ver. 2.0  
25 <210> SEQ ID NO: 1  
26 <211> LENGTH: 1525  
27 <212> TYPE: PRT  
28 <213> ORGANISM: Thermus aquaticus  
30 <220> FEATURE:  
31 <221> NAME/KEY: SITE  
32 <222> LOCATION: (1247)  
33 <223> OTHER INFORMATION: Any amino acid can be at this position  
35 <400> SEQUENCE: 1  
36 Met Lys Lys Glu Val Arg Lys Val Arg Ile Ala Leu Ala Ser Pro Glu  
37    1               5               10               15  
39 Lys Ile Arg Ser Trp Ser Tyr Gly Glu Val Glu Lys Pro Glu Thr Ile  
40    20               25               30  
42 Asn Tyr Arg Thr Leu Lys Pro Glu Arg Asp Gly Leu Phe Asp Glu Arg  
43    35               40               45  
45 Ile Phe Gly Pro Ile Lys Asp Tyr Glu Cys Ala Cys Gly Lys Tyr Lys  
46    50               55               60  
48 Arg Gln Arg Phe Glu Gly Lys Val Cys Glu Arg Cys Gly Val Glu Val  
49    65               70               75               80  
51 Thr Arg Ser Ile Val Arg Arg Tyr Arg Met Gly His Ile Glu Leu Ala  
52    85               90               95  
54 Thr Pro Ala Ala His Ile Trp Phe Val Lys Asp Val Pro Ser Lys Ile  
55    100              105              110  
57 Gly Thr Leu Leu Asp Leu Phe Ala Thr Glu Leu Glu Gln Val Leu Tyr  
58    115              120              125  
60 Phe Asn Lys Tyr Ile Val Leu Asp Pro Lys Gly Ala Val Leu Asp Gly  
61    130              135              140  
63 Val Pro Val Glu Lys Arg Gln Leu Leu Thr Asp Glu Glu Tyr Arg Glu  
64    145              150              155              160  
66 Leu Arg Tyr Gly Lys Gln Glu Thr Tyr Pro Leu Pro Ala Gly Val Asp  
67    165              170              175  
69 Ala Leu Val Lys Asp Gly Glu Glu Val Val Lys Gly Gln Glu Leu Ala  
70    180              185              190

ENTERED

RECEIVED  
APR -4 2001  
OIPE/JCWS

RAW SEQUENCE LISTING DATE: 03/29/2001  
 PATENT APPLICATION: US/09/782,714 TIME: 09:59:32

Input Set : N:\Crf3\RULE60\09782714.txt  
 Output Set: N:\CRF3\03292001\I782714.raw

72 Pro Gly Val Val Ser Arg Met Asp Gly Val Gly Ser Leu Pro Leu Pro  
 73 195 200 205  
 75 Arg Arg Val Arg Val Asp Tyr Leu Arg Lys Glu Arg Ala Ala Leu Arg  
 76 210 215 220  
 78 Ile Pro Leu Ser Ala Trp Val Glu Lys Glu Pro Tyr Arg Pro Gly Glu  
 79 225 230 235 240  
 81 Val Leu Ala Glu Leu Ser Glu Pro Tyr Leu Phe Arg Ala Glu Glu Ser  
 82 245 250 255  
 84 Gly Val Val Glu Leu Lys Asp Leu Ala Glu Gly His Leu Ile Tyr Leu  
 85 260 265 270  
 87 Arg Gln Glu Glu Val Val Ala Arg Tyr Phe Leu Pro Ala Gly Met  
 88 275 280 285  
 90 Thr Pro Leu Val Val Glu Gly Glu Ile Val Glu Val Gly Gln Pro Leu  
 91 290 295 300  
 93 Ala Glu Gly Lys Gly Leu Leu Arg Leu Pro Arg His Met Thr Ala Lys  
 94 305 310 315 320  
 96 Glu Val Glu Ala Glu Glu Gly Asp Ser Val His Leu Thr Leu Phe  
 97 325 330 335  
 99 Leu Glu Trp Thr Glu Pro Lys Asp Tyr Lys Val Ala Pro His Met Asn  
 100 340 345 350  
 102 Val Ile Val Pro Glu Gly Ala Lys Val Gln Ala Glu Lys Ile Val  
 103 355 360 365  
 105 Ala Ala Ile Asp Pro Glu Glu Val Ile Ala Gln Ala Glu Gly Val  
 106 370 375 380  
 108 Val His Leu His Glu Pro Ala Ser Ile Leu Val Val Lys Ala Arg Val  
 109 385 390 395 400  
 111 Tyr Pro Phe Glu Asp Asp Val Glu Val Thr Thr Gly Asp Arg Val Ala  
 112 405 410 415  
 114 Pro Gly Asp Val Leu Ala Asp Gly Gly Lys Val Lys Ser Glu Ile Tyr  
 115 420 425 430  
 117 Gly Arg Val Glu Val Asp Leu Val Arg Asn Val Val Arg Val Val Glu  
 118 435 440 445  
 120 Ser Tyr Asp Ile Asp Ala Arg Met Gly Ala Glu Ala Ile Gln Glu Leu  
 121 450 455 460  
 123 Leu Lys Glu Leu Asp Leu Glu Lys Leu Glu Arg Glu Leu Leu Glu Glu  
 124 465 470 475 480  
 126 Met Lys His Pro Ser Arg Ala Arg Arg Ala Lys Ala Arg Lys Arg Leu  
 127 485 490 495  
 129 Glu Val Val Arg Ala Phe Leu Asp Ser Gly Asn Arg Pro Glu Trp Met  
 130 500 505 510  
 132 Ile Leu Glu Ala Val Pro Val Leu Pro Pro Asp Leu Arg Pro Met Val  
 133 515 520 525  
 135 Gln Val Asp Gly Gly Arg Phe Ala Thr Ser Asp Leu Asn Asp Leu Tyr  
 136 530 535 540  
 138 Arg Arg Leu Ile Asn Arg Asn Asn Arg Leu Lys Lys Leu Leu Ala Gln  
 139 545 550 555 560  
 141 Gly Ala Pro Glu Ile Ile Ile Arg Asn Glu Lys Arg Met Leu Gln Glu  
 142 565 570 575  
 144 Ala Val Asp Ala Val Ile Asp Asn Gly Arg Arg Gly Ser Pro Val Thr

RAW SEQUENCE LISTING DATE: 03/29/2001  
 PATENT APPLICATION: US/09/782,714 TIME: 09:59:32

Input Set : N:\Crf3\RULE60\09782714.txt  
 Output Set: N:\CRF3\03292001\I782714.raw

|     |   |     |     |     |
|-----|---|-----|-----|-----|
| 145 | 580   | 585 | 590 |     |
| 147 | Asn Pro Gly Ser Glu Arg Pro Leu Arg Ser Leu Thr Asp Ile Leu Ser |     |     |     |
| 148 | 595   | 600 | 605 |     |
| 150 | Gly Lys Gln Gly Arg Phe Arg Gln Asn Leu Leu Gly Lys Arg Val Asp |     |     |     |
| 151 | 610   | 615 | 620 |     |
| 153 | Tyr Ser Gly Arg Ser Val Ile Val Val Gly Pro Gln Leu Lys Leu His |     |     |     |
| 154 | 625   | 630 | 635 | 640 |
| 156 | Gln Cys Gly Leu Pro Lys Arg Met Ala Leu Glu Leu Phe Lys Pro Phe |     |     |     |
| 157 | 645   | 650 | 655 |     |
| 159 | Leu Leu Lys Lys Met Glu Glu Lys Ala Phe Ala Pro Asn Val Lys Ala |     |     |     |
| 160 | 660   | 665 | 670 |     |
| 162 | Ala Arg Arg Met Leu Glu Arg Gln Arg Asp Ile Lys Asp Glu Val Trp |     |     |     |
| 163 | 675   | 680 | 685 |     |
| 165 | Asp Ala Leu Glu Glu Val Ile His Gly Lys Val Val Leu Leu Asn Arg |     |     |     |
| 166 | 690   | 695 | 700 |     |
| 168 | Ala Pro Thr Leu His Arg Leu Gly Ile Gln Ala Phe Gln Pro Val Leu |     |     |     |
| 169 | 705   | 710 | 715 | 720 |
| 171 | Val Glu Gly Gln Ser Ile Gln Leu His Pro Leu Val Cys Glu Ala Phe |     |     |     |
| 172 | 725   | 730 | 735 |     |
| 174 | Asn Ala Asp Phe Asp Gly Asp Gln Met Ala Val His Val Pro Leu Ser |     |     |     |
| 175 | 740   | 745 | 750 |     |
| 177 | Ser Phe Ala Gln Ala Glu Ala Arg Ile Gln Met Leu Ser Ala His Asn |     |     |     |
| 178 | 755   | 760 | 765 |     |
| 180 | Leu Leu Ser Pro Ala Ser Gly Glu Pro Leu Ala Lys Pro Ser Arg Asp |     |     |     |
| 181 | 770   | 775 | 780 |     |
| 183 | Ile Ile Leu Gly Leu Tyr Tyr Ile Thr Gln Val Arg Lys Glu Lys Lys |     |     |     |
| 184 | 785   | 790 | 795 | 800 |
| 186 | Gly Ala Gly Met Ala Phe Ala Thr Pro Glu Glu Ala Leu Ala Ala Tyr |     |     |     |
| 187 | 805   | 810 | 815 |     |
| 189 | Glu Arg Gly Glu Val Ala Leu Asn Ala Pro Ile Val Val Ala Gly Arg |     |     |     |
| 190 | 820   | 825 | 830 |     |
| 192 | Glu Thr Ser Val Gly Arg Leu Lys Phe Val Phe Ala Asn Pro Asp Glu |     |     |     |
| 193 | 835   | 840 | 845 |     |
| 195 | Ala Leu Leu Ala Val Ala His Gly Leu Leu Asp Leu Gln Asp Val Val |     |     |     |
| 196 | 850   | 855 | 860 |     |
| 198 | Thr Val Arg Tyr Leu Gly Arg Arg Leu Glu Thr Asn Pro Gly Arg Ile |     |     |     |
| 199 | 865   | 870 | 875 | 880 |
| 201 | Leu Phe Ala Arg Ile Val Gly Glu Ala Val Gly Asp Glu Lys Val Ala |     |     |     |
| 202 | 885   | 890 | 895 |     |
| 204 | Gln Glu Leu Ile Gln Met Asp Val Pro Gln Glu Lys Asn Ser Leu Lys |     |     |     |
| 205 | 900   | 905 | 910 |     |
| 207 | Asp Leu Val Tyr Gln Ala Phe Leu Arg Leu Gly Met Glu Lys Thr Ala |     |     |     |
| 208 | 915   | 920 | 925 |     |
| 210 | Arg Leu Leu Asp Ala Leu Lys Tyr Tyr Gly Phe Thr Leu Ser Thr Thr |     |     |     |
| 211 | 930   | 935 | 940 |     |
| 213 | Ser Gly Ile Ile Thr Ile Gly Ile Asp Asp Ala Val Ile Pro Glu Glu |     |     |     |
| 214 | 945   | 950 | 955 | 960 |
| 216 | Lys Gln Arg Tyr Leu Glu Ala Asp Arg Lys Leu Arg Gln Ile Glu     |     |     |     |
| 217 | 965   | 970 | 975 |     |

RAW SEQUENCE LISTING DATE: 03/29/2001  
 PATENT APPLICATION: US/09/782,714 TIME: 09:59:32

Input Set : N:\Crf3\RULE60\09782714.txt  
 Output Set: N:\CRF3\03292001\I782714.raw

219 Gln Ala Tyr Glu Met Gly Phe Leu Thr Asp Arg Glu Arg Tyr Asp Gln  
 220 980 985 990  
 222 Val Ile Gln Leu Trp Thr Glu Thr Thr Glu Lys Val Thr Gln Ala Val  
 223 995 1000 1005  
 225 Phe Asn Asn Phe Glu Glu Asn Tyr Pro Phe Asn Pro Leu Tyr Val Met  
 226 1010 1015 1020  
 228 Ala Gln Ser Gly Ala Arg Gly Asn Pro Gln Gln Ile Arg Gln Leu Cys  
 229 1025 1030 1035 1040  
 231 Gly Met Arg Gly Leu Met Gln Lys Pro Ser Gly Glu Thr Phe Glu Val  
 232 1045 1050 1055  
 234 Pro Val Arg Ser Ser Phe Arg Glu Gly Leu Thr Val Leu Glu Tyr Phe  
 235 1060 1065 1070  
 237 Ile Ser Ser His Gly Ala Arg Lys Gly Gly Ala Asp Thr Ala Leu Arg  
 238 1075 1080 1085  
 240 Thr Ala Asp Ser Gly Tyr Leu Thr Arg Lys Leu Val Asp Val Ala His  
 241 1090 1095 1100  
 243 Glu Ile Val Val Arg Glu Ala Asp Cys Gly Thr Thr Lys Tyr Ile Ser  
 244 1105 1110 1115 1120  
 246 Val Pro Leu Phe Gln Met Asp Glu Val Thr Arg Thr Leu Arg Leu Arg  
 247 1125 1130 1135  
 249 Lys Arg Ser Asp Ile Glu Ser Gly Leu Tyr Gly Arg Val Leu Ala Arg  
 250 1140 1145 1150  
 252 Glu Val Glu Ala Leu Gly Arg Arg Leu Glu Glu Gly Arg Tyr Leu Ser  
 253 1155 1160 1165  
 255 Leu Glu Asp Val His Phe Leu Ile Lys Ala Ala Glu Ala Gly Glu Val  
 256 1170 1175 1180  
 258 Arg Glu Val Pro Val Arg Ser Pro Leu Thr Cys Gln Thr Arg Tyr Gly  
 259 1185 1190 1195 1200  
 261 Val Cys Gln Lys Cys Tyr Gly Tyr Asp Leu Ser Met Ala Arg Pro Val  
 262 1205 1210 1215  
 264 Ser Ile Gly Glu Ala Val Gly Val Val Ala Ala Glu Ser Ile Gly Glu  
 265 1220 1225 1230  
 W--> 267 Pro Gly Thr Gln Leu Thr Met Arg Thr Phe His Thr Gly Gly Xaa Ala  
 268 1235 1240 1245  
 270 Val Gly Thr Asp Ile Thr Gln Gly Leu Pro Arg Val Ile Glu Leu Phe  
 271 1250 1255 1260  
 273 Glu Ala Arg Arg Pro Lys Ala Lys Ala Val Ile Ser Glu Ile Asp Gly  
 274 1265 1270 1275 1280  
 276 Val Val Arg Ile Glu Glu Gly Glu Asp Arg Leu Ser Val Phe Val Glu  
 277 1285 1290 1295  
 279 Ser Glu Gly Phe Ser Lys Glu Tyr Lys Leu Pro Lys Asp Ala Arg Leu  
 280 1300 1305 1310  
 282 Leu Val Lys Asp Gly Asp Tyr Val Glu Ala Gly Gln Pro Leu Thr Arg  
 283 1315 1320 1325  
 285 Gly Ala Ile Asp Pro His Gln Leu Leu Glu Ala Lys Gly Pro Glu Ala  
 286 1330 1335 1340  
 288 Val Glu Arg Tyr Leu Val Asp Glu Ile Gln Lys Val Tyr Arg Ala Gln  
 289 1345 1350 1355 1360  
 291 Gly Val Lys Leu His Asp Lys His Ile Glu Ile Val Val Arg Gln Met

RAW SEQUENCE LISTING DATE: 03/29/2001  
 PATENT APPLICATION: US/09/782,714 TIME: 09:59:32

Input Set : N:\Crf3\RULE60\09782714.txt  
 Output Set: N:\CRF3\03292001\I782714.raw

292 1365 1370 1375  
 294 Leu Lys Tyr Val Glu Val Thr Asp Pro Gly Asp Ser Pro Leu Leu Glu  
 295 1380 1385 1390  
 297 Gly Gln Val Leu Glu Lys Trp Asp Val Glu Ala Leu Asn Glu Arg Leu  
 298 1395 1400 1405  
 300 Ile Ala Glu Gly Lys Val Pro Val Ala Trp Lys Pro Leu Leu Met Gly  
 301 1410 1415 1420  
 303 Val Thr Lys Ser Ala Leu Ser Thr Lys Ser Trp Leu Ser Ala Ala Ser  
 304 1425 1430 1435 1440  
 306 Phe Gln Asn Thr Thr His Val Leu Thr Glu Ala Ala Ile Ala Gly Lys  
 307 1445 1450 1455  
 309 Lys Asp Glu Leu Ile Gly Leu Lys Glu Asn Val Ile Leu Gly Arg Leu  
 310 1460 1465 1470  
 312 Ile Pro Ala Gly Thr Gly Ser Asp Phe Val Arg Phe Thr Gln Val Val  
 313 1475 1480 1485  
 315 Asp Gln Arg Thr Leu Lys Ala Ile Glu Glu Ala Arg Lys Glu Ala Val  
 316 1490 1495 1500  
 318 Glu Ala Lys Glu Lys Glu Ala Pro Arg Arg Pro Val Arg Arg Glu Gln  
 319 1505 1510 1515 1520  
 321 Pro Gly Lys Gly Leu  
 322 1525  
 325 <210> SEQ ID NO: 2  
 326 <211> LENGTH: 1119  
 327 <212> TYPE: PRT  
 328 <213> ORGANISM: Thermus aquaticus  
 330 <220> FEATURE:  
 331 <221> NAME/KEY: SITE  
 332 <222> LOCATION: (695)..(696)  
 333 <223> OTHER INFORMATION: Any amino acids can be at these two positions.  
 335 <400> SEQUENCE: 2  
 336 Met Lys Ile Lys Arg Phe Gly Arg Ile Arg Glu Val Ile Pro Leu Pro  
 337 1 5 10 15  
 339 Pro Leu Thr Glu Ile Gln Val Glu Ser Tyr Lys Lys Ala Leu Gln Ala  
 340 20 25 30  
 342 Asp Val Pro Pro Glu Lys Arg Glu Asn Val Gly Ile Gln Ala Ala Phe  
 343 35 40 45  
 345 Lys Glu Thr Phe Pro Ile Glu Gly Asp Lys Gly Lys Gly Leu  
 346 50 55 60  
 348 Val Leu Asp Phe Leu Glu Tyr Arg Ile Gly Asp Pro Pro Phe Ser Gln  
 349 65 70 75 80  
 351 Asp Glu Cys Arg Glu Lys Asp Leu Thr Tyr Gln Ala Pro Leu Tyr Ala  
 352 85 90 95  
 354 Arg Leu Gln Leu Ile His Lys Asp Thr Gly Leu Ile Lys Glu Asp Glu  
 355 100 105 110  
 357 Val Phe Leu Gly His Leu Pro Leu Met Thr Glu Asp Gly Ser Phe Ile  
 358 115 120 125  
 360 Ile Asn Gly Ala Asp Arg Val Ile Val Ser Gln Ile His Arg Ser Pro  
 361 130 135 140  
 363 Gly Val Tyr Phe Thr Pro Asp Pro Ala Arg Pro Gly Arg Tyr Ile Ala

VERIFICATION SUMMARY DATE: 03/29/2001  
PATENT APPLICATION: US/09/782,714 TIME: 09:59:33

Input Set : N:\Crf3\RULE60\09782714.txt  
Output Set: N:\CRF3\03292001\I782714.raw

L: 267 M: 341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L: 465 M: 341 W: (46) "n" or "Xaa" used, for SEQ ID#:2